Compositionality prediction of Multiword Expressions

Silvio Ricardo Cordeiro
Aline Villavicencio
Carlos Ramisch

Outline

- Introduction
- Compositionality dataset
- Compositionality prediction
- Conclusions

Introduction

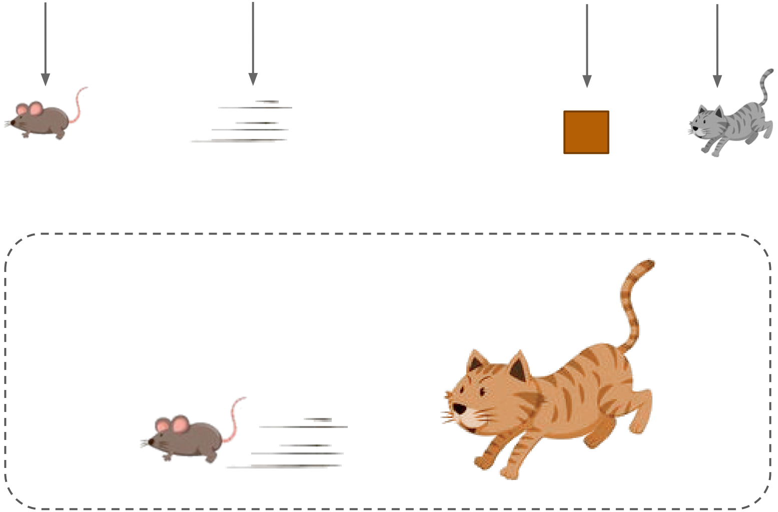
Computational Semantics

- We are interested in semantics.
  - Representing the meaning of words and sentences.

- Computational semantics has applications in:
  - Machine translation.
  - Information extraction.
  - Text simplification.
  - Question answering.
  - …

Principle of compositionality

- The meaning of the **whole** comes from the meaning of the **parts**.
- *“The mouse is running from the brown cat”*



Distributional Semantics

- At the word-level: Distributional Hypothesis
  - *"You shall know a word by the company it keeps"* — Firth, 1957

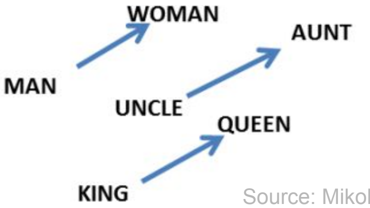
- Consider the word **fish**:

… and **fish** that swim between submerged branches need to …
… wondering: where do **fish** learn to swim? ...
... you may see some **fish** as you walk along the river …
… if the **fish** are swimming upstream, they will …
… we swim in a three-dimensional world, among parrot **fish** and …
… as you swim along, you can see different kinds of **fish** …

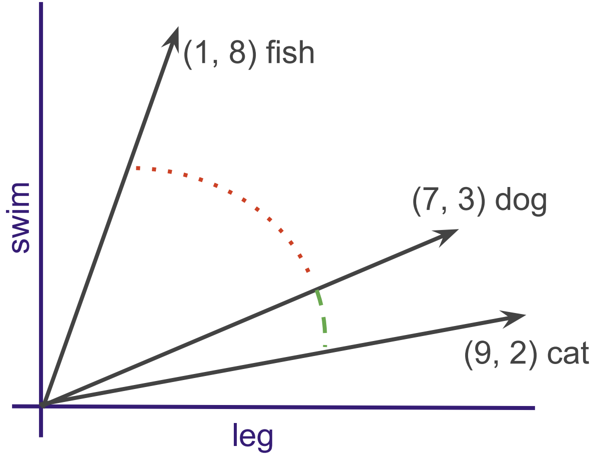
Distributional Semantics

- Distributional Semantic Models (DSMs):
    - Each word has a representation in $\mathbb{R}^n$
    - Source: words in context.

- Properties:
    - Similar concepts are near each other.
    - Vector arithmetic (e.g. for analogy tasks).



Source: Mikolov [2013]



Distributional Semantics

- Weakness: Multiword Expressions (MWEs).
- MWEs can range from compositional to idiomatic:
  - *climate change* > … > *milk tooth* > … > *hot dog* > *cloud nine*
- Non-compositional cases need special treatment.
  - Our goal: automatically detect the level of *compositionality*.
- Assume this hypothesis:
  - MWE is compositional ⇔ MWE is similar to the sum of its meanings
    - e.g. $v(climate\_change) \approx v(climate) + v(change)$.

Not a hot dog.

Nominal compounds

- We focus on a type of MWE known as nominal compounds.
  - More specifically: **noun-noun** and **adjective-noun** compounds.

snow storm

modifier (noun) **head**

escada rolante

head modifier (adjective)

Main contributions

- To construct & analyze compositionality **datasets**.
- To provide a **pipeline** for compositionality prediction.
  - Including a token-based MWE identifier.
- To **evaluate** DSM models & parameters for compositionality prediction.

Compositionality datasets

Compositionality datasets

- MWEs and their *compositionality*
  - *Numerical judgments* through crowdsourcing
  - Useful for evaluating compositionality prediction

- Reddy et al. [2011]
  - 90 English nominal compounds
  - ~15 mechanical turkers annotate each compound
  - Each compound is given a score between 0 and 5

- Farahmand et al. [2014]
  - 1042 English nominal compounds
  - 4 experts giving each compound a score of 0 or 1

MWE			
nut_case	1	1	1
labour_union	5	5	4
engine_room	5	5	5
milk_tooth	2	3	3
...	...	...	...

Compositionality datasets

- We adapt the methodology of Reddy and Farahmand:
    - Multiple languages: English, French, and Portuguese
        - 180 compounds for each language
    - For each compound:
        - ~15 annotators (Mechanical Turk)
        - Annotators must provide at least 2 synonyms
        - Requested compositionality judgments between 0 and 5
            - Judgments for <u>head</u>, <u>modifier</u> and for the <u>compound</u> as a whole

Dataset collection questionnaire

1. Read the following expression:

pocket book

2. Read the following sentences containing the expression *pocket book*:

- *All of these are at good prices to suit your* ***pocket book****.*
- *He gave me some Spanish books and a* ***pocket book*** *and diary.*
- *She had written down the date in her* ***pocket book*** *of the day when she dispatched it.*

3. Type in 2 to 3 expressions that are equivalent to *pocket book*: Required

4. In your opinion, is a *pocket book* always literally a *book*?

NO 0 1 2 3 4 5 YES

5. In your opinion, is the meaning of a *pocket book* always literally related to *pocket*?

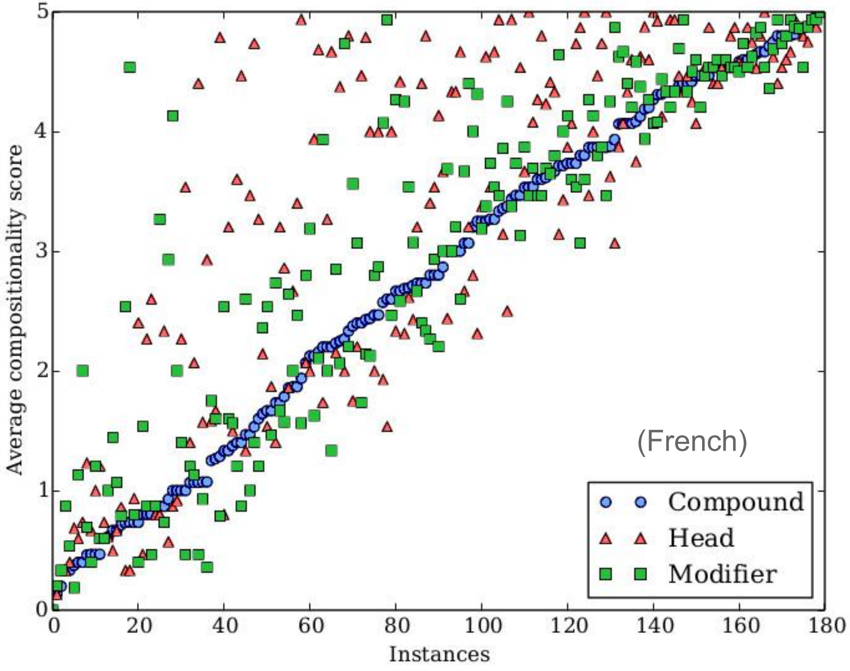
NO 0 1 2 3 4 5 YES

6. Given your previous replies, would you say that a *pocket book* is always literally a *book* which is related to *pocket*?

NO 0 1 2 3 4 5 YES

No — it is weird to imagine a *book* which is related to *pocket*, even if the meaning is understandable

Compositionality scores

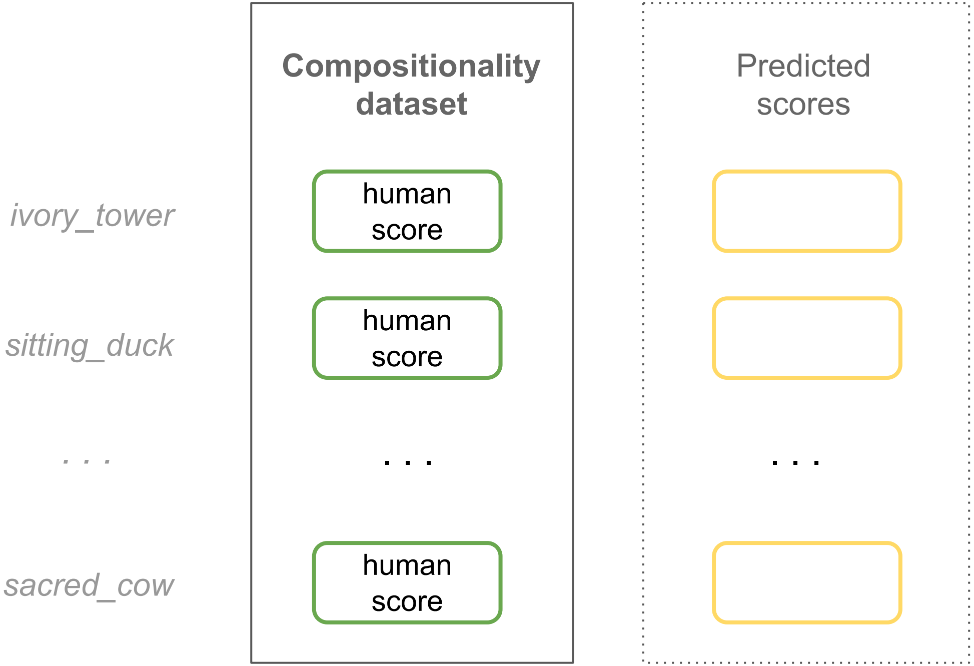


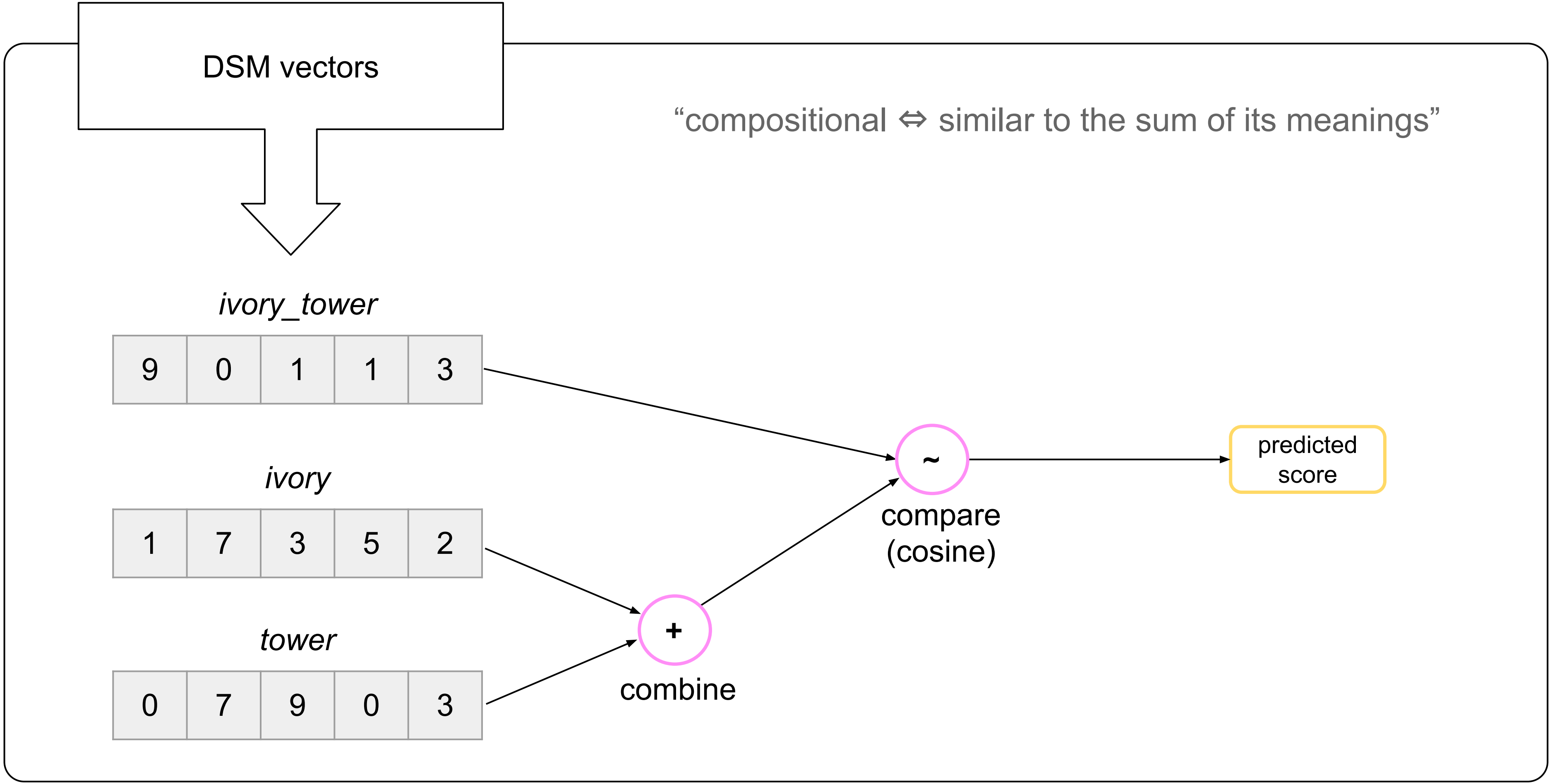
- All 3 datasets:
  - Balanced in compositionality.
  - Head/mod have a pattern.

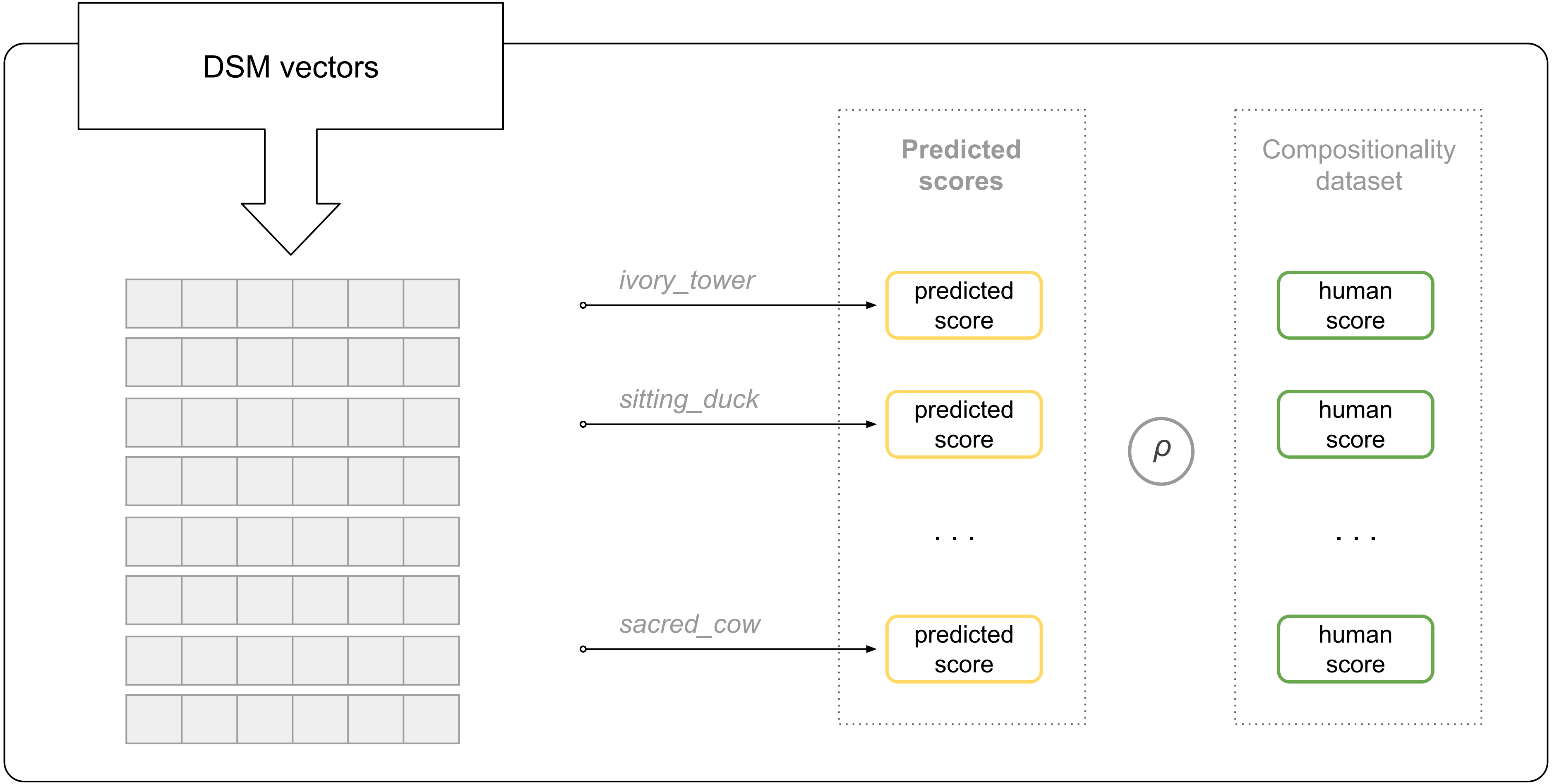
Ramisch, Cordeiro, Zilio, Idiart, Villavicencio, Wilkens.
How Naked is the Naked Truth? A Multilingual Lexicon of Nominal Compound Compositionality. In: **ACL 2016 (short paper)**. Qualis: **A1**.

Compositionality prediction

Compositionality prediction




DSM vectors
“compositional ⇔ similar to the sum of its meanings”
ivory_tower
9 0 1 1 3
ivory
1 7 3 5 2
tower
0 7 9 0 3
+
combine
~
compare
(cosine)
predicted
score


DSM vectors
ivory_tower
sitting_duck
sacred_cow
Predicted scores
predicted score
predicted score
. . .
predicted score
ρ
Compositionality dataset
human score
human score
. . .
human score

Spearman Correlation

Ordered Human scores

1 hot_dog
2 milk_tooth
3 sacred_cow
4 middle_school
5 climate_change

Ordered Prediction A

1 hot_dog
2 milk_tooth
3 sacred_cow
4 middle_school
5 climate_change

$\rho = +1$

Ordered Prediction B

5 climate_change
4 middle_school
3 sacred_cow
2 milk_tooth
1 hot_dog

$\rho = -1$

Ordered Prediction C

1 hot_dog
2 milk_tooth
5 climate_change
3 sacred_cow
4 middle_school

$\rho = +0.7$

Ordered Prediction D

4 middle_school
1 hot_dog
3 sacred_cow
5 climate_change
2 milk_tooth

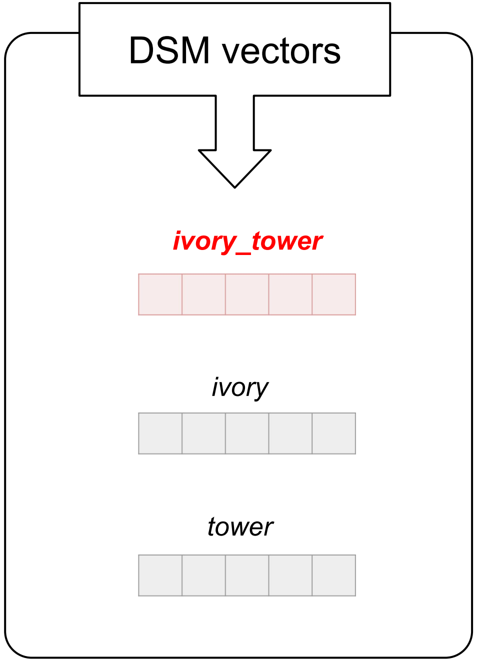
$\rho = 0$

Compositionality prediction pipeline

- We have implemented a pipeline as part of the mwetoolkit:
    - Read MWEs & DSM vectors.
    - For each MWE:
        - Combine its components and compare against the MWE itself.
        - The comparison results in a predicted compositionality score.
    - Calculate correlation between prediction and human scores.

Cordeiro, Ramisch, Villavicencio. *mwetoolkit+sem: Integrating Word Embeddings in the mwetoolkit for Semantic MWE Processing*. In: **LREC 2016**. Qualis: **A2**.

MWE identification



- DSM vectors **must include MWEs**.
- We have implemented a MWE identifier.
  - Works on multiple corpus formats.
  - Good F_1 for noun compounds.

Cordeiro, Ramisch, Villavicencio. *UFRGS&LIF: Rule-Based MWE Identification and Predominant-Supersense Tagging*. In: **SemEval 2016**. Qualis: **B4**.

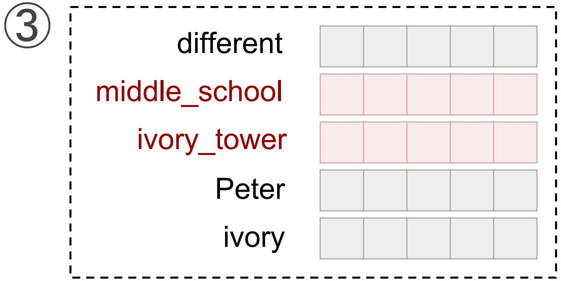
Full pipeline

① When we were in middle school, Peter was a very different guy. He needs to get down from his ivory tower and come eat a hot dog with us.

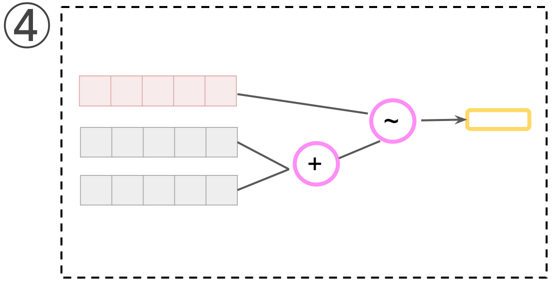
Input corpus

② When we were in middle_school, Peter was a very different guy. He needs to get down from his ivory_tower and come eat a hot_dog with us.

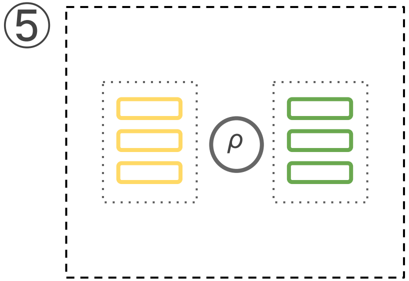
MWE identification



DSM creation



Score prediction



Evaluation

DSMs & parameters

- Our next goal is to investigate DSMs & parameters:

DSMs

- ➔ **PPMI-TopK**: global contexts [Salehi et al., 2015]
- ➔ **PPMI-thresh**: local context threshold
- ➔ **PPMI-SVD**: dimensionality reduction [Dinu et al, 2013]
- ➔ **glove**: dimensionality reduction [Pennington et al., 2014]
- ➔ **w2v** (word2vec): neural networks [Mikolov, 2013]

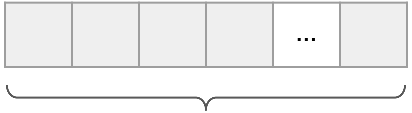
DSMs & parameters

We know that, when **fish** swim upstream, they often ...

Context window size: 1, 4, 8.

Preprocessing

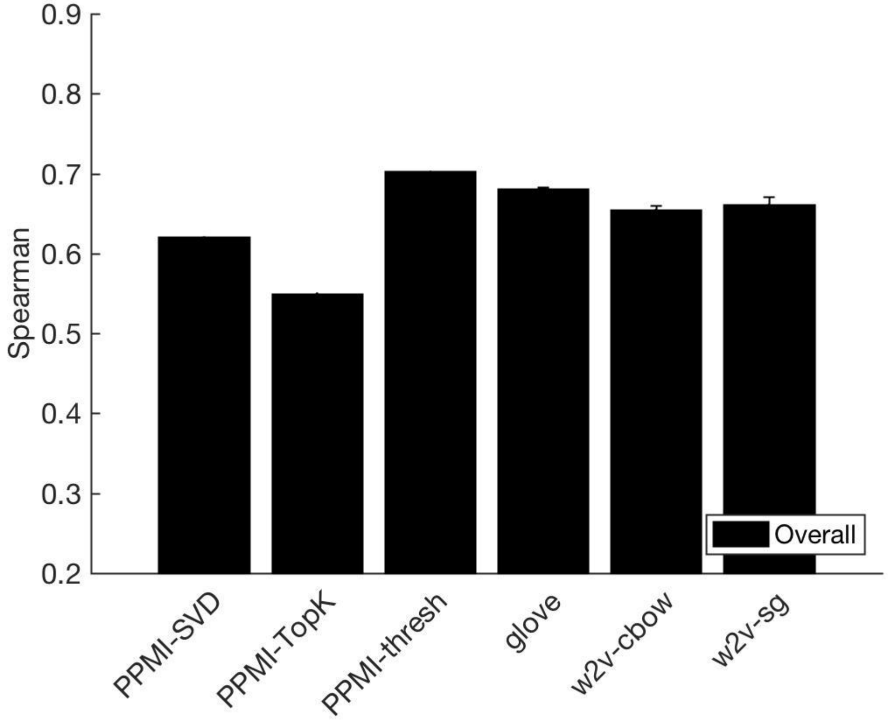
- ➔ lemma (e.g. **walk**)
- ➔ lemmaPOS (e.g. **walk/VERB**)
- ➔ surface (e.g. **walks**)
- ➔ surface+ (with stopwords)



Dimensions: 250, 500, 750.

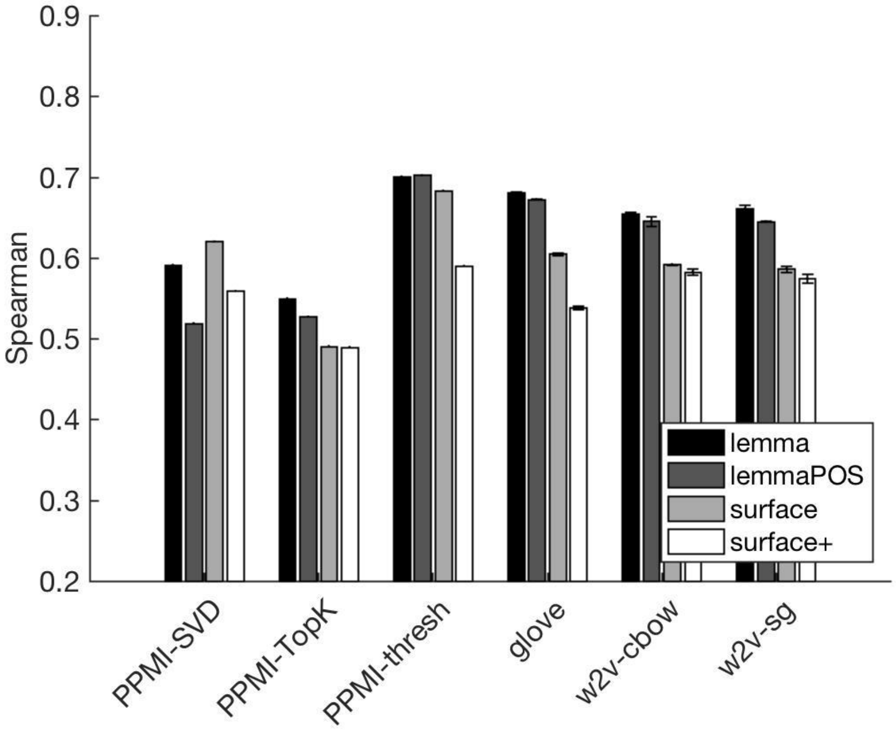
- Total of 816 models.
- We present the results for our datasets.

Highest results for French dataset



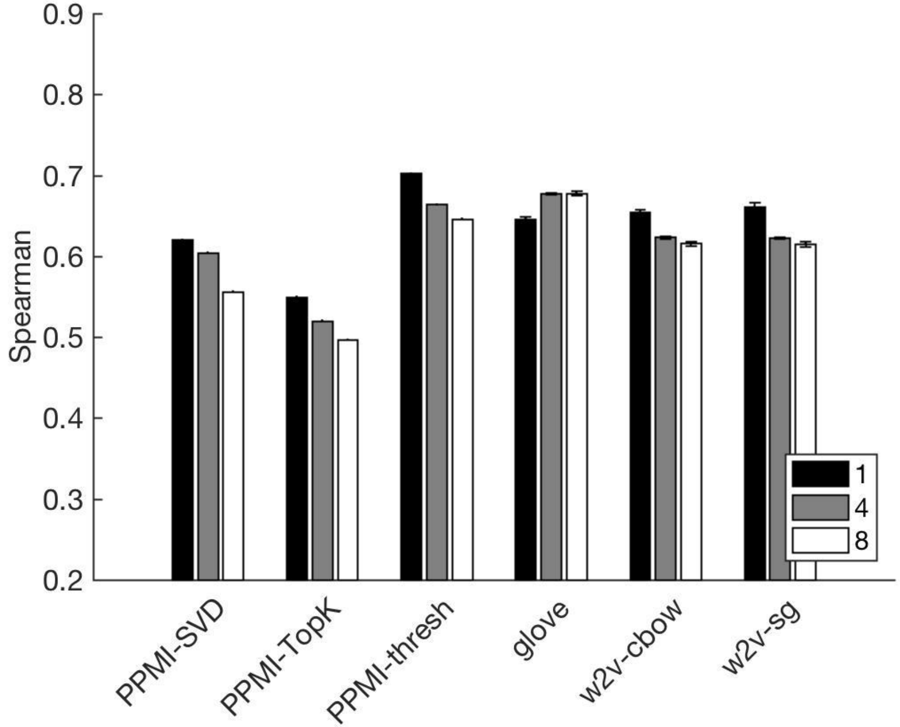
- SVD: ...
- Global contexts: ✗
- Classical model: ✔

Highest results for French dataset



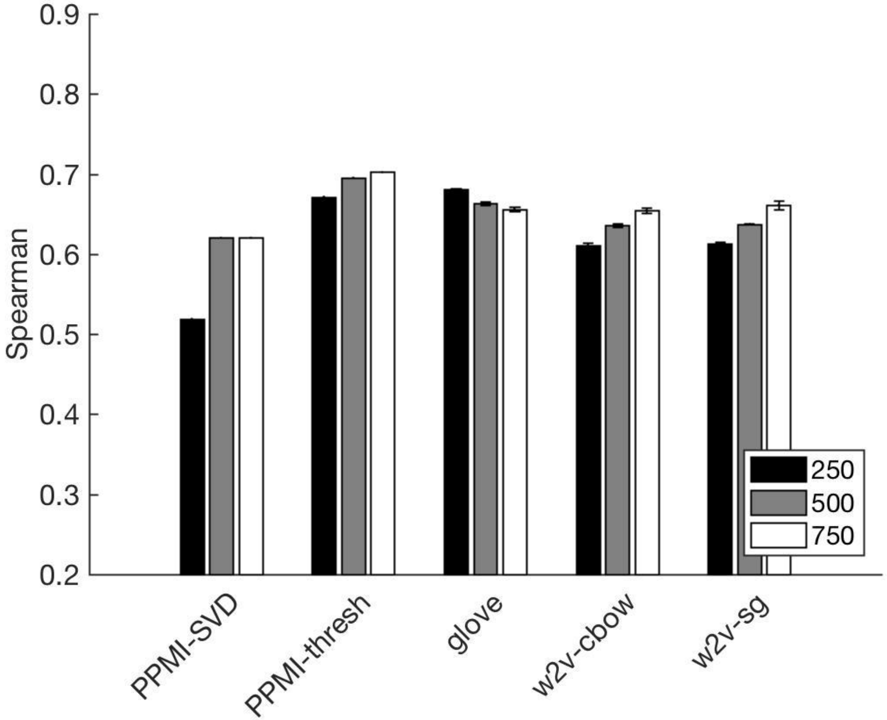
- Stopword removal: ✔
- Lemmatization: ✔
- POS-tagging: ...

Highest results for French dataset



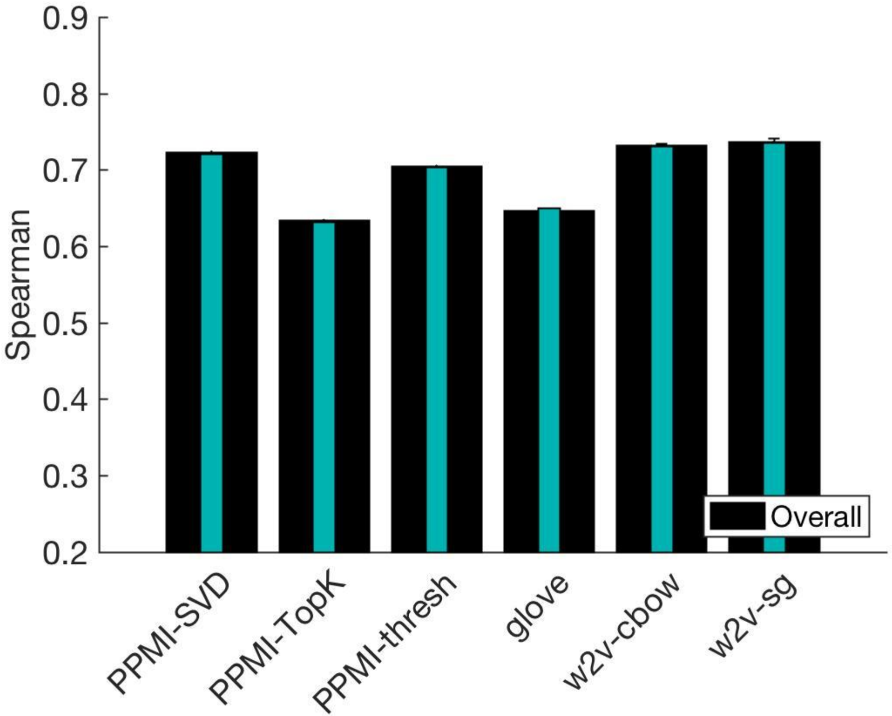
$w_1 > w_4 > w_8$

Highest results for French dataset



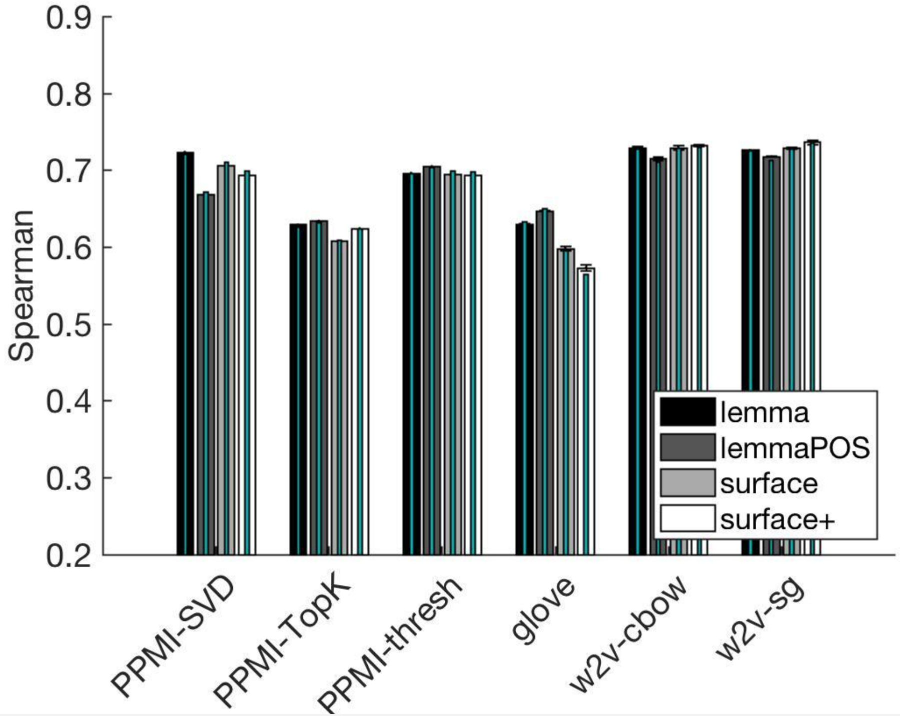
$d_{250} < d_{500} < d_{750}$

Highest results for English dataset



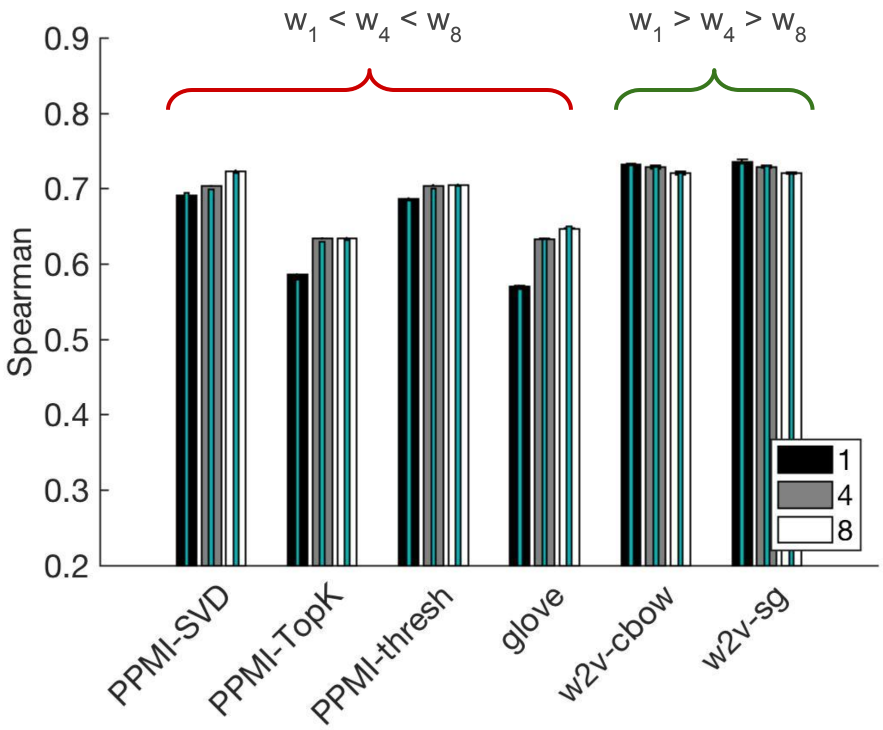
- SVD: ✔
- Glove: ✗

Highest results for English dataset

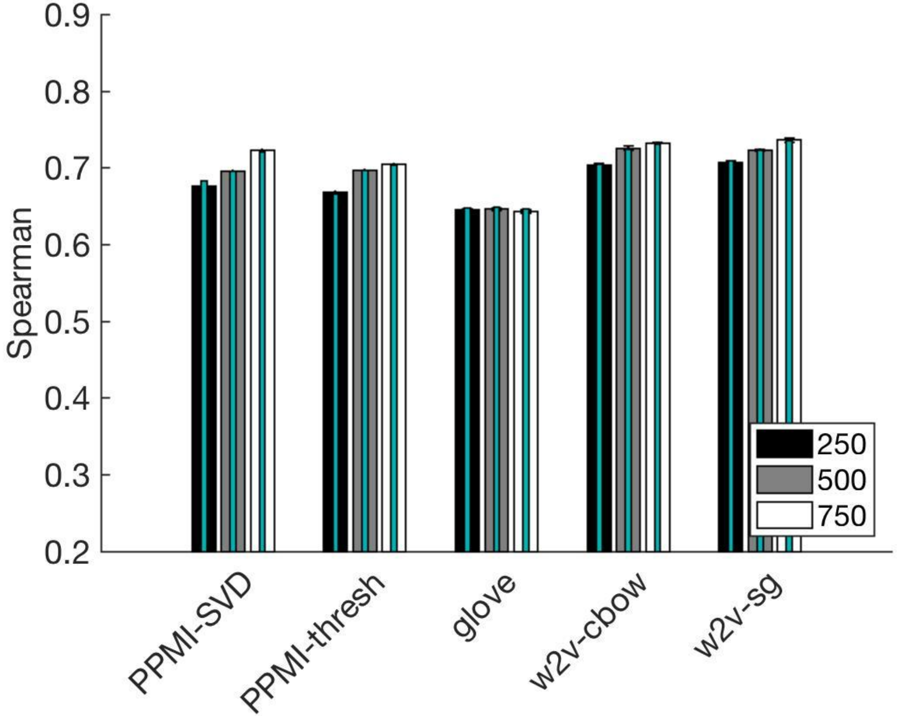


- Lemmatization: ✔
- Surface-forms: ✔
- Stopword removal: ?

Highest results for English dataset



Highest results for English dataset



$d_{250} < d_{500} < d_{750}$ ✔

Comparing with State of the Art

- Dataset from Reddy et al [2011]:

Model & Parameters	Spearman ρ
Reddy et al [2011]	.71
Salehi et al [2015]	.80
Best *w2v* (*sg, WF=surface, D=750, W=1*)	**.82** / .80
Best *PPMI* (*thresh, WF=surface, D=750, W=8*)	**.80** / .80

- Dataset from Farahmand et al [2015]:

Model & Parameters	Best-F_1
Yazdani et al [2015]	.49
Best *w2v* (*sg, WF=lemma, D=500, W=1*)	**.51** / .47
Best *PPMI* (*svd, WF=lemma, D=750, W=4*)	**.52** / .45

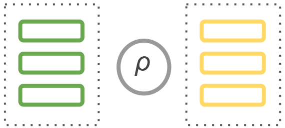
Cordeiro, Ramisch, Idiart, Villavicencio.
Predicting the Compositionality of Nominal Compounds:
Giving Word Embeddings a Hard Time.
In: **ACL 2016 (long paper)**. Qualis: **A1**.

Conclusions

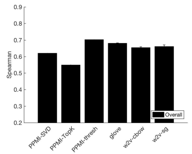
Conclusions

- Constructed 3 compositionality **datasets**.
  - Also evaluated statistical properties and the impact of filtering.

- Built a compositionality prediction **pipeline**.
  - Corpus → Corpus+MWEs → DSM vectors → predict & evaluate.



- Performed extensive **evaluation** of DSMs & parameters.
  - Classical model as good as neural networks.
  - Higher number of dimensions often better.
  - Lemmas better for French, not impactful for English.
  - POS-tags are often unhelpful.



Publications

- Cordeiro, Ramisch, Idiart, Villavicencio. *Predicting the Compositionality of Nominal Compounds: Giving Word Embeddings a Hard Time*. In: **ACL 2016 (long paper)**. Qualis: A1.
- Ramisch, Cordeiro, Zilio, Idiart, Villavicencio, Wilkens. *How Naked is the Naked Truth? A Multilingual Lexicon of Nominal Compound Compositionality*. In: **ACL 2016 (short paper)**. Qualis: A1.
- Cordeiro, Ramisch, Villavicencio. *mwetoolkit+sem: Integrating Word Embeddings in the mwetoolkit for Semantic MWE Processing*. In: **LREC 2016**. Qualis: A2.
- Ramisch, Cordeiro, Villavicencio. *Filtering and Measuring the Intrinsic Quality of Human Compositionality Judgments*. In: **MWE 2016**. Qualis: B3.
- Zilio, Wilkens, Möllmann, Wehrli, Cordeiro, Villavicencio. *Joining forces for multiword expression identification.* In: **PROPOR 2016**. Qualis: B3.
- Cordeiro, Ramisch, Villavicencio. *UFRGS&LIF: Rule-Based MWE Identification and Predominant-Supersense Tagging*. In: **SemEval 2016**. Qualis: B4.

Publications

- Cordeiro, Ramisch, Villavicencio. *Token-based MWE Identification Strategies in the mwetoolkit*. In: **PARSEME 2015**.
- Cordeiro, Ramisch, Villavicencio. *Nominal Compound Compositionality: A Multilingual Lexicon and Predictive Model*. In: **PARSEME 2016**.
- Cordeiro, Ramisch, Villavicencio. *MWE-aware corpus processing with the mwetoolkit and word embeddings*. In: **W-PROPOR 2016**.

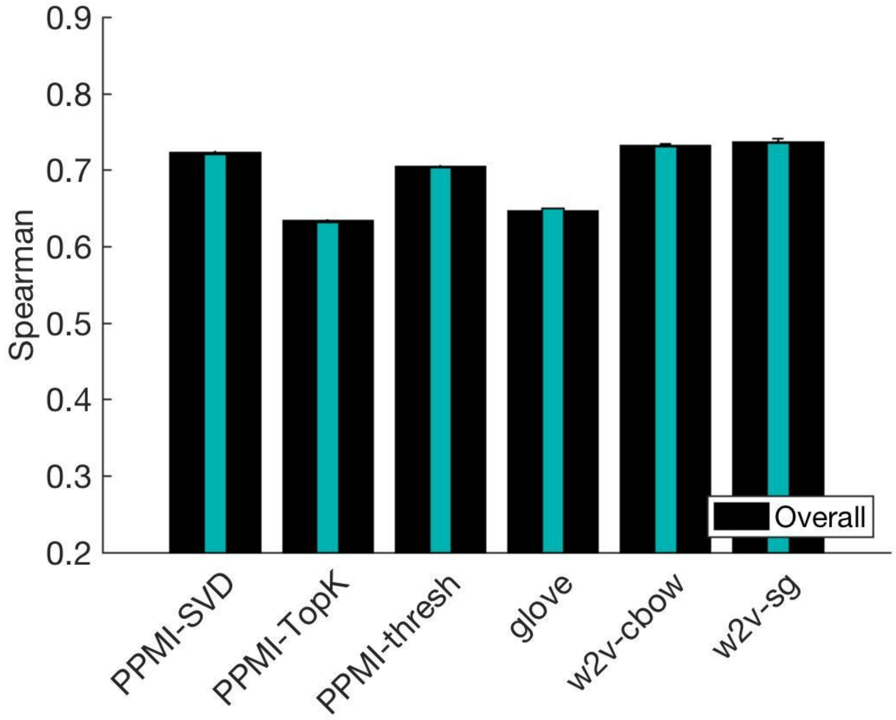
Compositionality prediction of Multiword Expressions

Thank you!

Silvio Ricardo Cordeiro
Aline Villavicencio
Carlos Ramisch

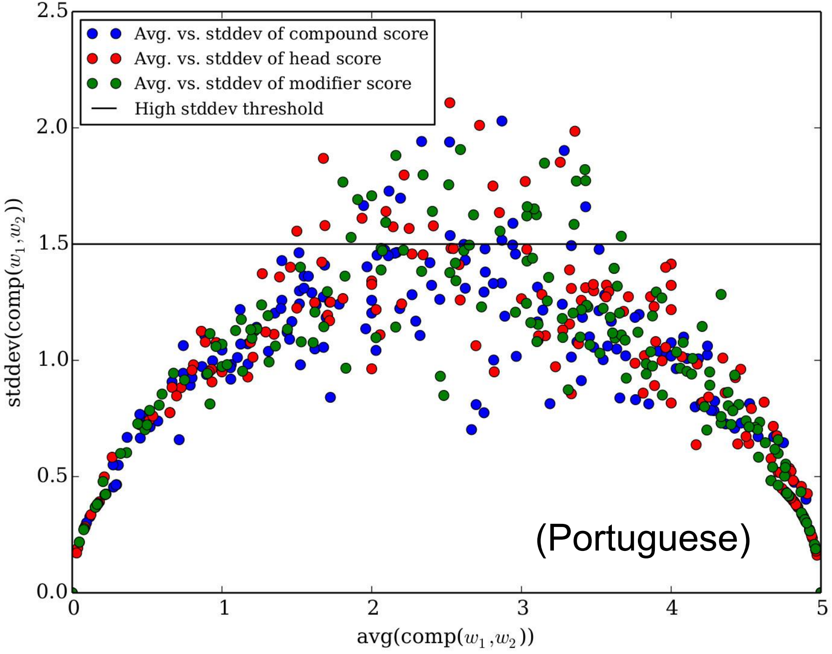
(Extra slides)

Highest results for English: strict vs loose

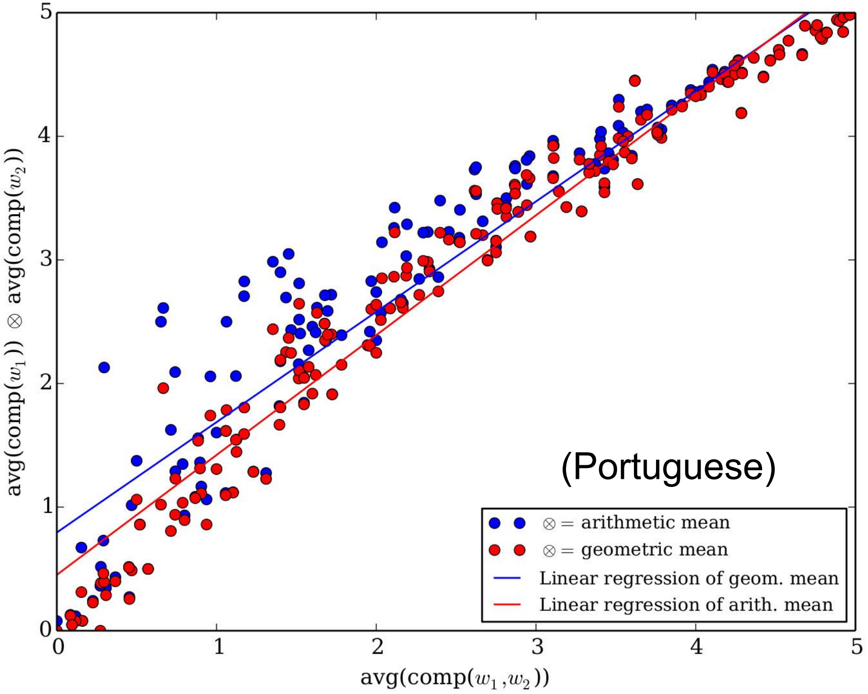


Missing data:
strict (smaller dataset)
loose (fallback)

Compounds vs difficulty of annotation



Approximating whole-compound judgment



Problem #1 meets #2

- MWEs can be polysemic:
  - “I just ate a delicious piece of cake” → compositional
  - “The test was a piece of cake” → non-compositional

Compositionality prediction

(Work in progress)

Work in progress

- Currently investigating other configurations:
  - Portuguese language.
  - Impact of corpus size.
  - Composing scores from smaller corpora (ensemble).
  - Analyzing the influence of *head* vs *mod* in score.
  - …

- Goal: submit a paper to Computational Linguistics.

Planned research

Polysemy

Planned research: Polysemy

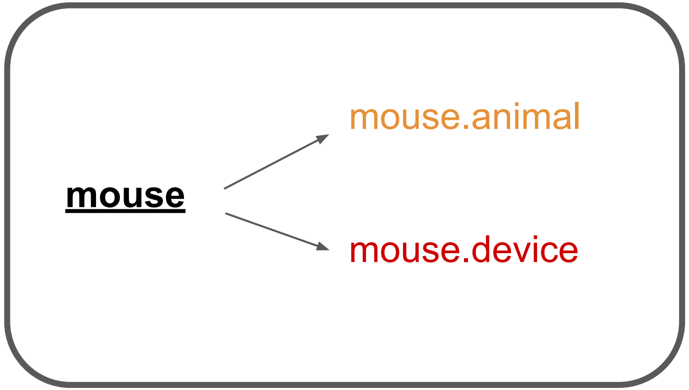
mouse → mouse.animal
mouse → mouse.device

Word-sense induction of types

… the lion let the **mouse** go … → mouse.animal
… playing a cat and **mouse** game … → mouse.animal
… an owl catching a **mouse** … → mouse.animal
… when a **mouse** click occurs … → mouse.device
... pressing the **mouse** selects the … → mouse.device
… using a **mouse** and keyboard … → mouse.device

Word-sense disambiguation of tokens

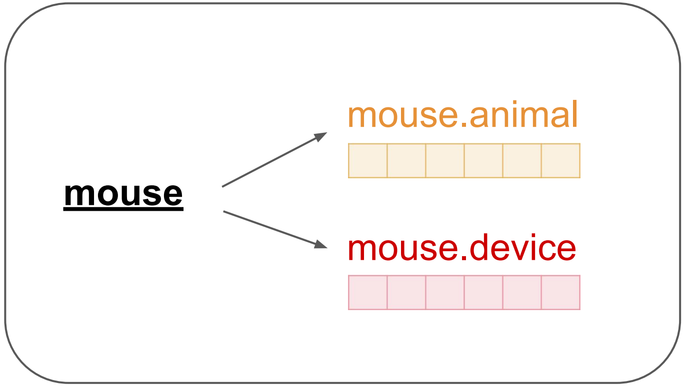
Planned research: Polysemy



Word-sense induction of types

- We will focus on sense induction
  - Pre-requisite for good disambiguation.
  - Current solutions ignore MWE.
  - We have an idea.

Planned research: Polysemy

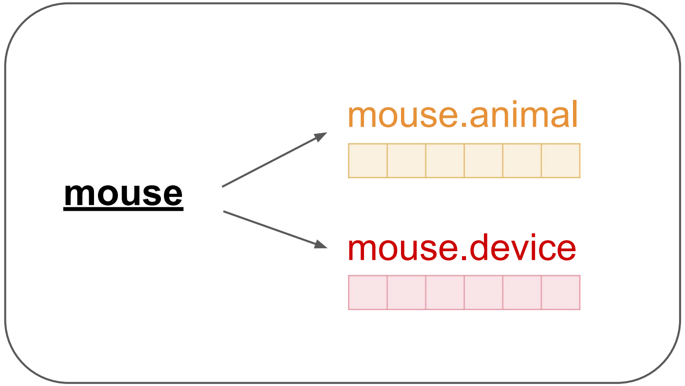


Word-sense induction of types

… the lion let the **mouse** go … → n.animal
… playing a cat and **mouse** game … → n.animal
… an owl catching a **mouse** … → n.animal
… when a **mouse** click occurs … → n.artifact
... pressing the **mouse** selects the … → n.artifact
… using a **mouse** and keyboard … → n.artifact

Annotated corpus from SemEval 2016 task 10

Planned research: Polysemy



Word-sense induction of types

Distributional Hypothesis: use contexts!

… the lion let the **mouse** go … → n.animal

… playing a cat and **mouse** game … → n.animal

… an owl catching a **mouse** … → n.animal

… when a **mouse** click occurs … → n.artifact

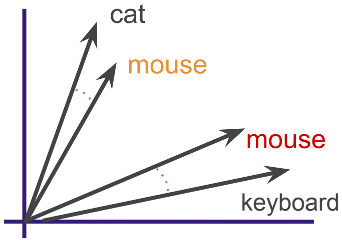
... pressing the **mouse** selects the … → n.artifact

… using a **mouse** and keyboard … → n.artifact

Annotated corpus from SemEval 2016 task 10

Planned research: Polysemy

- Evaluation:
  - Word similarity task (e.g. SimLex-999 dataset)
    - Idea: Use most similar sense when comparing for sinonimy
    - e.g.: *mouse* ≈ *cat*; *mouse* ≈ *keyboard*; *cat* $\not\approx$ *keyboard*



  - Compositionality task (e.g. our compositionality datasets)
    - Idea: Use most similar sense when testing compositionality
    - e.g.: *mouse trap* and *mouse click* would be both compositional